Comments on Impact Evaluation of the Farmer Training and Development Activity in Honduras
Markus Goldstein

This report seeks to provide an impact evaluation of one of the two objectives of the MCA-Honduras Program, i.e. to "increase the productivity and business skills of farmers who operate small and medium size farms and their employees."   The report discusses a failed randomized design and then uses an ex-post propensity score matching design.

The propensity score estimates indicate that a) there was a significant increase in expenditure on basic grain production and, more relevant to the activities at hand, that there was an increase in the net income from horticultural crops.   There were no significant increases in household income or expenditure (all estimates relative to the control group).

As the report is currently written, I do not find these results plausible.   The main issue is captured in different appendices.   First, as the authors indicate in Annex III the critical assumption underlying propensity score matching is that there is not likely to be selection on unobservables.   Unfortunately, the clear frustration of the evaluation team with the selection process of program implementers indicated throughout the report, and the discussion in Annex I, makes me wonder about the validity of this assumption.   For example as the discussion in Annex I indicates: " new farmers were not really selected, but 'self-identified'…Technicians report that these new recruits are both *very motivated* and *more aware* of the level of effort involved in implementing the program." and "If during their visits to aldeas, they notice a plot where a farmer has done a *reasonable* job planting, the technician will try and recruit him into the program…" and "according to several technicians, the trend towards self-identification has *increased the quality* of program farmers" (emphases added)

From a program point of view, this is good news – these are probably the farmers for whom the program is likely to have the greatest impact.   However, from a quasi-experimental evaluation point of view this is vexing – variables that capture attributes such as motivation, ability, and quality of farmers are next to impossible to capture.   What is more puzzling is that, while the authors are clearly technically competent, and there was clearly some thoughtful working on unpacking the selection process, there is no clear discussion of these selection problems in the main text.   A more clear discussion of these problems, and what it spells for the reliability of the estimation, needs to be front and center in the discussion of the analysis. The lack of this discussion of what is clearly a central problem for the chosen method makes me concerned about the overall validity of the results (not least of which because I do not see a clear way to resolve this problem).

Given this fundamental problem, we unfortunately cannot conclude whether or not this project was successful in achieving its objectives.   Perhaps, if the data set were rich enough, or other characteristics were used to identify a potential control group (e.g. the discussion on selection into the program seems to indicate that capacity constraints may have created rationing at some point) we could get an estimate that, while still flawed, was more plausible.

The experience with this evaluation could prove instructive to the MCC on the pitfalls of randomized design, but this would require further explanations which seem to be beyond the scope of the paper as now written. In particular, it is clear that this was a train wreck of a randomization.   A deeper post-mortem could provide lessons on how to avoid this in the future.   A couple of points on this (and these are not meant to reflect on the shortcomings of the evaluation team – I am in no position to judge this,

but rather to highlight some potential (and perhaps painful) experiences that could provide lessons with further frank reflection:

- The initial design was elegant, thoughtful and well laid out
- It wasn't clear to me why the evaluation activities commenced two years after the start of the program (page 3)
- There seems to have been a communication gap between Fintrac and the evaluation team, with the evaluation team chasing Fintrac around trying to ex-post recover the revisions to the selection process. How could this have been avoided? Would have embedding someone within Fintrac ameliorated this? Moreover, there were deviations from originally stated criteria for things such as aldea-level screening by Fintrac.   While MCA and NORC engaged Fintrac on this (ex-post of course), would a stronger enforcement mechanism have preserved things?
- It still isn't clear to me (and perhaps not to the evaluation team) why Fintrac rejected so many of the pre-screened farmers. How could this have been avoided?
- Was poverty reduction an underlying goal of the program? If so, could a stricter adherence to wider program criteria (and not some version of cream skimming as seems to have happened) saved the evaluation and kept the program more in line with fundamental objectives? If so, who could have done this?